Presidential Commission
for the Study of Bioethical Issues

TRANSCRIPT

Latanya Sweeney, Ph.D.
Visiting Professor and Scholar, Computer Science
Director, Data Privacy Lab
Harvard University

Sonia Suter, M.S., J.D.
Professor of Law
George Washington University

Meeting 10, Opening Remarks and Session 1
August 1, 2012
Washington, DC

DR. GUTMANN: Good morning, everybody. I'm Amy Gutmann. I'm president of the University of Pennsylvania and chair of the Presidential Commission for the Study of Bioethical Issues. On behalf of my vice chair Jim Wagner who is president of Emory University and myself I welcome you all to this our tenth meeting.

Before we continue and in order for us to continue officially I want to recognize our designated federal officer, Dr. Lisa Lee. Lisa, would you please stand up? Lisa is the executive director of our Commission. Thank you.

So we have two full days ahead of us, today and tomorrow morning. We're going to continue our discussion of the ethics of whole genome sequencing. We began our work earlier this year, last year actually and we've had many stakeholders, experts, members of the public present and enrich our deliberations. We also have reached out to 18 federal agencies to learn about their relevant policies and practices and they have been very forthcoming in their responses.

And I'm really pleased to say we've received extensive and thoughtful public comment in response to our request for information that's published in the Federal Register this past March. And we have circulated those comments among all the Commission members and we will take them into account in our final report.

Today as in our past meetings we will hear expert presentations. We will then transition to the sessions that really will enable the Commission to deliberate in public about our recommendations, to discuss what we are tentatively at this point but by the end of this meeting we will have some more developed sense of the Commission's recommendations to President Obama.

And after we conclude tomorrow's discussion section we will turn to another ongoing project which is the Commission's ethical review of pediatric medical countermeasures, the research on them and the ethics of them.

So I'd like to take a moment before we begin to explain how we will take comments from the audience. And I really encourage people who have questions to do this which is there are cards outside at the registration desk. You write your question and comment down. Give it to any member of the staff. They will deliver it up here to Jim and me and time permitting, and we hope there will be time, we will read and respond to your questions.

Would the members of the Commission staff please stand up so everyone knows who? And they all have name tags. There you go. And we brought some cards in here. So if you want cards they're easily accessible. Hillary, why don't you -- there we go. Okay.

That's all I have to say before we begin, but I'd like to give Jim Wagner, our vice chair, an opportunity to welcome everyone.

DR. WAGNER: Thank you, Amy, and good morning to you by the way and good morning to all. Like you, Amy, I have and we all have really benefitted from the input from expert testimony during our prior deliberations. We're grateful for public and federal agency comment on this subject and continue to be challenged by that fundamental charge, you know, that we got from the White House which was to discern what constitutes ethically appropriate and sufficient regulatory policy and practice to ensure that all society is able to maximize the benefits to be gained by the advancement of medical science and

technology, in this case this morning whole genome sequencing while at the same time minimizing risk to society and to individuals, especially those most vulnerable.

I think it's helpful every now and then to, at least it is for me, to be reminded of that charge because it is a charge to help chart a path for benefit and progress, recognizing that doing so will necessarily incur some level of risk. Our job is to discern and offer opinion on what constitutes acceptable risk and how to ensure that that level of risk is not exceeded.

So it's in that spirit that I welcome the commissioners this morning. I welcome our expert who will be with us, the staff, welcome to you folks, and the public who will participate and thank everyone for their assistance in this discernment process.

DR. GUTMANN: Terrific, thank you. So we're beginning our morning with a discussion of how technology is changing views of privacy. And it is the case that privacy while a continuing value has changed culturally as to where the lines that are considered private and public are over time. And technology certainly has contributed its fair share to those shifting lines.

We will hear from two speakers and then we'll open the session for questions and discussion. And I'm going to ask both of our speakers to come up. Do we have our -- is Latanya here? Welcome. And one reminder to you and all other speakers including Commission members, this button, make sure it's on when you're speaking and when you're not speaking turn it off. But if you leave it on I think there won't be a problem but make sure it's on when you are speaking.

So my pleasure, first, to welcome Dr. Latanya Sweeney who is a visiting professor and a visiting scholar at Harvard University. She is also the director and founder

of Harvard's Data Privacy Lab where she focuses on data identifiability and privacy. And those are twin poles of what we as a Commission are -- part of what we are focusing on.

Dr. Sweeney has testified before the Privacy and Integrity Advisory Committee of the Department of Homeland Security and the European Union. Her work has appeared in hundreds of news articles and numerous academic papers, and she has received awards for her work from the American Psychiatric Association, the American Medical Informatics Association and the Blue Cross Blue Shield Association.

Welcome, Dr. Sweeney.

DR. SWEENEY: Thank you, Dr. Gutmann, it's a pleasure to be here, and members of the President's Commission for the Study of Bioethics. Thank you for giving me an opportunity to talk with you about using technology to save privacy.

There's no doubt that technology does shape our expectations of privacy. In fact, technology has shaped much of how we are likely to talk about privacy today and that's not surprising. But what I do hope to propose to you that you might is a bit of a surprise is that we can craft new technology to shape how society might think of our privacy as we go forward.

And I use the term "technology" in this context to talk about an integration, sort of a bundle of technical capability, economics and policy. And they work together as a unit.

For example, the iPod was not the first MP3 player, but it was the first to integrate digital content easily, thereby tackling related economic and copyright issues. So I would say MP3 players provide technical know-how but the iPod is for our purposes a

technology.

In the United States where we have no comprehensive privacy law our experiences shape our collective expectations of privacy. In recent decades technology has been one of the greatest forces to shape our experiences and by its design our privacy expectations.

Google, Facebook and others offer good examples in data. People give intimate personal information about themselves freely in exchange for useful services. My mobile phone company records my text messages, voice messages, phone numbers and GPS locations. My hospital may keep my biospecimens indefinitely.

Last year PricewaterhouseCoopers estimated that the sharing of medical records beyond the care of the patient was a $2 billion market. The result on society is a belief that privacy is dead. There's just too much information out there.

On the other hand, once collected these data, my data, your data become closely guarded private assets of these companies. Usually I cannot get a copy of my own information nor can scientific researchers access the information for many worthy purposes.

So is the future of privacy to be determined by data market forces involving these privately held silos? The individuals who are the subject of the data have no say, yet individuals may personally bear the consequences.

I said that to sort of level-set the conversation because what I'd like to describe are some new living labs in which we help people assemble and control copies of their own information to improve their own lives and to provide information responsibly to research.

Before I jump into that though I would like to say the breadth of work that I've worked on is pretty wide and so I'll just only introduce a couple of these labs and then of course in conversation I'd be welcome to go elsewhere.

Let me just say a little bit about how I started and the vision. As a professor of computer science I have trained hundreds of young minds. Most recently I realized that these young computer scientists and computer science entrepreneurs are actually policymakers, able to make architectural decisions about technology that dictate real-world practice and establish societal norms. Yet we have never trained them for this role and as a consequence the privacy impact of their decisions is often ill-conceived and unintentional. Privacy considerations are an afterthought if a thought at all. So how then can we use technology, this bundle of technical know-how, economics and policy, to enhance privacy?

The vision is simple. We want society to enjoy the privacy and utility, and with the current lineup of technology we often falsely believe that society must choose between privacy and utility. Our vision is to deploy new technologies that help society enjoy both privacy and utility.

And I've had a lot of success in this area. It includes ways of assessing re-identification risk in data, methods for sharing data under the HIPAA statistician provision and so forth.

But today I'm going to talk about two new living labs at various stages of operation at launch at Harvard, these new living labs which are technology approaches at shaping the future.

Let me just -- first I'll list some of the living labs that are out there. MyDataCan is a lab in which members of the public are invited to assemble and control their own data. Data to Science, a project in which people donate data to science after death. Open Consent or the Personal Genome Project which is at the medical school where people share data freely on the internet. theDataMap is a public display of secondary data-sharing arrangements. And PrivaMix, a means of computing information across silos without moving the data. I'm only going to talk in the interest of time about the first two of those.

So MyDataCan.org is a living lab that allows members of the public to collect, assemble and distribute their own personal data across data silos including health information without a fee and optionally elect to participate in activities that use their data to improve the quality of their lives. This approach combines data that's otherwise trapped in silos giving the individual the most complete copy of information about themselves. And it adds transparency and knowledge to data-sharing arrangements.

Having so much personal data about an individual creates an economic ecosystem for third party apps. So the platform has an app development environment similar to what you see on smartphones. The idea is that others who have already starting coming forward will display data, show data, use your GPS locations, complying with other information to help you make decisions. And that has already begun to take place.

Let me talk a moment now about the other project and then I'll stop. The DatatoScience.org, this is an activity on the MyDataCan platform. It helps a person donate his personal data to scientific research after his death. Just as a person can donate his organs

to science through organ donor programs the DatatoScience project allows a person to donate data collected about him to science. And these data may be tremendously helpful for retrospective research, promising to improve and understand ourselves and how to live life better.

The system allows a person to register their information, sign an agreement that survives past their death and gives stewardship. And the system upon their demise will actually go and compile the information and release it.

One thing about both Data to Science and MyDataCan is they have an additional privacy guarantee that even the holders of the data cannot view the data. And the Data to Science project, you can't view the data until their demise even if it's already collected. And in the MyDataCan you can't view the data at all, it's sort of doubly encrypted where Harvard has one key but the other key is maintained by the individual. And so even if we were to receive a court order all we could do is use our key and provide encrypted results.

So in concluding we believe these kinds of living labs will provide benefits and utility to society and begin to establish new norms for privacy expectations.

DR. GUTMANN: Thank you very much. Next I'm really delighted that we will hear from Professor Sonia Suter. Professor Suter is a law professor at George Washington University Law School teaching courses in genetics and the law and law and medicine.

Professor Suter's scholarship focuses on legal and ethical issues in medicine and genetics, including privacy protections of genetic information, genetic

exceptionalism and DNA forensics.

Professor Suter has advised and worked with policymakers on issues related to genetics and bioethics. And prior to law school she obtained her master's of science in human genetics and worked as a genetic counselor.

Welcome, Dr. Suter.

MS. SUTER: Thank you very much for having me here today. It's a real honor to participate in these discussions.

I was asked to talk about how to reconcile the public and privacy interests in genomic research as well as to talk about how technology is influencing views of privacy, a tall order in 10 minutes. So I'm going to focus on a few specific aspects of this.

At the outset I do want to emphasize that although we're focusing on genetic information here I don't think genetic information is exceptional, but much of it is sensitive like a great deal of other medical information. So I want to sort of have that caveat.

I want to talk a little bit about the privacy interests. The primary privacy interest that I think we have in this context is the ability to control our genetic information, to control who has access to it, what we call informational privacy. And I think that this also includes the right or ability to decide what information we get about ourselves.

I also think that physical privacy interests are implicated as well as decisional privacy interests. But I want to also add that I think when we think about the sharing of genetic information and samples that we should think about a trust-based notion of privacy. Generally when we make decisions to share this information we're doing so in

the context of relationships of trust, not arms' length transactions, and therefore there are obligations of trust on the part of the recipient because of the vulnerabilities that this imposes on the source.

I want to talk a little bit about the basis of informational privacy interests because I saw this came up in some of your discussions pretty frequently and I want to talk about that also in the context of how courts seem to be thinking about privacy with emerging technologies.

So, a while back the Supreme Court did recognize that there was a constitutional interest in informational privacy in Whalen v. Roe when they noted that we have a constitutional interest in avoiding disclosure of personal information.

But the Court didn't see this is an absolute interest. It found it constitutional to require that prescriptions for level 2 drugs include patient information and be disclosed to the Department of Health. This was justified as a reasonable exercise of the police powers, a kind of balancing of the public good against privacy interests.

But I think it's important to note that the Court also said that in doing this there was a corollary obligation to do two things, to make sure that the information was used for public purposes and to avoid unwanted disclosures, the trust notion that I was talking about.

The other area where courts have been thinking about privacy is in Fourth Amendment jurisprudence which is a narrower notion of privacy but I think relevant here. And it shows some of the challenges of thinking about privacy when it becomes increasingly difficult to keep our information secret.

In my view, unfortunately, the few courts that have looked at the constitutionality of the government's surreptitious searches for DNA have found those searches to be constitutional. And the real reason for this, the basis seems to be that we have no real reasonable expectation of privacy in abandoned materials generally. And the courts view these DNA searches as similar to other legitimate searches for identifying information that we leave behind in public.

But of course DNA is more than just identifying, it contains a great deal of personal and predictive information about health, personality, abilities and other traits. Moreover, we don't voluntarily abandon our DNA which I think should be relevant in our considerations of the privacy interests.

The Court was a little more protective of privacy in Ferguson when the Court recognized that if you voluntarily relinquish bodily fluids for prenatal care that doesn't mean you have an expectation that it will be searched for other purposes like searching for incriminating evidence of drug use.

So if we think about the relinquishment of biospecimens and samples in the context of relationships of trust, the doctor-patient relationship, the researcher-participant relationship, I think it's clear that people have fairly strong expectations of privacy and confidentiality. But in other contexts I think our expectations of privacy may be changing with technological developments.

Everybody has seen shows like CSI so pretty much everyone knows that the government or anyone who has the technology can get identifying information from samples left on a cup. With the growth of direct-to-consumer testing I think the people will

become increasingly aware of the ease of getting more than identifying information from samples. And in fact, it's fairly easy. It may not be accepted by the companies, but it's fairly easy to send a sample to a company for analysis of someone else, surreptitious analysis.

So we're at an interesting time where I think our law hasn't really sorted out what to do about the fact that we can leave all sorts of personal information about ourselves through our actions. Whether it's actions on the internet or simply walking down the streets and leaving cells with DNA information, or when we choose to relinquish samples or medical records for large-scale research.

By engaging in these activities I don't think that we have the expectation that we're waiving our privacy rights. And so I want to turn to one other Fourth Amendment case which is the recent decision United States v. Jones. And I think that Justices Alito's and Sotomayor's concurring opinions reflect two different kinds of views about privacy with changing technologies. This was the GPS tracking device case.

On the one hand Justice Alito says that "even if the public does not welcome the diminution of privacy that new technology entails they may eventually reconcile themselves to this development as inevitable." So one view is it's too late, we have no more privacy.

But Justice Sotomayor has a more complex view. She doubts that people would accept without complaint the warrantless disclosure to the government of a list of every website they had visited in a week, month, or year. And what I think is particularly interesting is the next statement.

"Whatever the societal expectations [these visits to websites] can attain constitutionally protected status only if our Fourth Amendment ceases to treat secrecy as a prerequisite for privacy."

And I think this is sort of what's at the heart of the issue here as we gather more data in the context of genetics research. With advancing technologies it's increasingly hard to keep secret our genetic information. There's more data-sharing, samples can be de-identified more easily than people may realize. But that doesn't mean that we don't have privacy interests here, it just means that we may need more explicit protections of those interests.

So one way to protect privacy is to give people control over access to their genetic information and samples. This is the notion of informational privacy that I just discussed. But to many, consent provisions -- of course it depends which ones exactly we have -- are at the heart of the conflict between privacy interests and the public good of research.

There are many concerns that if we give people too much control over their information and samples that will limit research in numerous ways. It might become impossible or impracticable, time-consuming, expensive, and there are a lot of worries about consent and selection bias.

So there's clearly a great public good in moving forward much of the genomics research. But how can we reconcile the privacy interests with this public good? The answer is it's not easy, but we do have to keep continually in mind that we have competing duties here both to privacy and to research.

And I think instead of a utilitarian approach or a deontological approach we need to recognize prima facie duties to both of these, and to recognize that neither one of them is absolute or unconditional. When the duties are in conflict as they are here we aren't going to be able to completely honor our prima facie duties to either and so we need to accept some limitations.

Now, the ideal kind of solution, and we may disagree about how we work that out, is to find a solution where we can maximize as much as possible the benefits of each and minimize the harms when we violate the other duties to some extent.

It's a somewhat abstract principle so let me just throw out one possible way of thinking about this. I'm not deeply wedded to this. There are some I still need to think through. But to fulfill the obligations of trust because we are after all asking the public to give up something for the public good we need to impose mandatory data security and information protection standards. This is the sort of notion that Whalen talked about when they acknowledged that the statute was constitutional.

Perhaps even require that the information really is used for the public good. That's a vague notion but some sort of promise to the people that when they're giving up a little bit of privacy rights that they're doing it for the public good since altruism motivates people to participate in research. We need rules and penalties to prohibit inappropriate re-identification of samples and to prohibit inappropriate uses of genetic information. And I think it's really important to have laws that prohibit the surreptitious kind of sampling and analysis that is allowed in most states, or isn't prohibited.

To acknowledge the needs for research while giving people some control

over their samples and their information but not full control we might think about having a kind of general consent for research with some options to opt out of future research and particular categories of research.

And then I also think it's really important to ensure that people have full understanding of the limitations of privacy protections in this area; that while there are efforts to maintain security of the data that it's not entirely failsafe. One of the greatest risks of participating in this research is the fact that there are privacy risks and people should be aware of that.

We might consider some waiver of consent in certain situations but there would have to be some pretty stringent conditions for that. Thank you.

DR. GUTMANN: Thank you very much. Let me begin with a question that just draws the two of your presentations together and then I'm going to open it up to members of the Commission and Vice Chair Wagner to ask any questions they have.

And thank you for two very lucid presentations. And you're both certainly working in areas that are directly relevant to our report.

So, Dr. Sweeney, you mentioned two organizations that collect data. And my question to you and it's a companion question to Dr. Suter is what kinds of consent do you ask for people who volunteer. These are voluntary sites of sharing data about your person. What's the form of consent that is gotten for this?

And then for Dr. Suter, you mentioned in conclusion that you think some general forms of consent would be adequate and yet a very general or blanket consent is not permissible under current policy for IRBs. IRBs will not agree to blanket consent forms.

So can you be more specific about how what you're recommending is different from -- if it is different from a blanket consent?

DR. SWEENEY: Yes. So, in the laundry list of projects that I listed the kinds of consents are open consent where you basically say when the person donates their data, like the Personal Genome Project is an example, they basically are saying not only do I not hold you to any privacy constraint, I hereby give you all the data and I agree to hold you harmless for any consequences. That would be open consent. Consent made is the general kind of consent model where the person says I give my data over to science and I'm consenting.

But the two projects that I listed to you are radically different. They do the exact opposite. They replace consent with a contract. And they bring into play the idea that the individual has assembled data that belongs to them, that's their copy and they have a sense of ownership over that copy of assembled data.

So we provide a mechanism to help them assemble the data to get the data out of the silos, assemble it, and this sort of richer data set is their own copy. It doesn't hold any issue about what happens to the other data in the other silos, only this copy.

And this copy is tightly controlled in the following way, that it's not controlled in that it's hard to share it. We make it easy to share it, but we share it in a way that the sharing is very transparent, limited to direct sharing. And in a lot of cases, in the cases of those apps the data never leaves the platform. The app gets delivered, it uses the data, tells them which of their data fields it's using and the data doesn't go outside. So I think that that would be the short version.

MS. SUTER: Okay, so the kind of consent -- I mean, there are I think nuances that one could create. But I think the idea is that we can't -- because of the impracticalities of trying to get informed consent for every kind of protocol that happens, and because the data may be used in ways that we can't fully anticipate in advance that I think that may be one of the compromises that we can make about research with samples.

It's not one I'm 100 percent happy with, but I think that trying to find something that honors both privacy and the value of research is going to require concessions on both sides. So I think what I had in mind was something where people would be informed about the possibilities of different kinds of research, about the possibilities of data-sharing depending on what the setup is for the collection of samples. And that people would have an option to opt out of particular categories. You could track that with the samples data or opt out of all future research.

But I think that informed consent, full informed consent of the kind that we imagine for IRBs is becoming increasingly difficult to achieve here. And I think that goes to autonomy interests and also to privacy, so I think there's sort of two things in play. We're focusing primarily on privacy but in the informational research I think the biggest risks are about data security.

And the most important thing is for people to understand what the limits are of privacy protections, that there will be efforts with some kind of system to secure the data, but that it isn't maybe as perfect a protection as we had once thought that it was. There may -- depending on the kind of research there may be other sorts of risks as well but we're not dealing with the same kind of risks that people face when they're dealing with

interventional research. So I think we may think about consent models in slightly different ways for that reason.

DR. GUTMANN: I thought your conclusion was at least gesturing or signaling to what I think is an important recognition both empirically and morally which is there may be ethical obligations of people who have received consent to use data that are not legally enforceable. So if you give consent, what you call open consent you're not expecting the people who you gave consent to to publicize the fact that they've -- that your whole genome sequencing on the front page of the Wall Street Journal or the New York Times suggests you may have propensity to a personality disorder or early onset Alzheimer's. I mean there are limits, ethical limits I would think that people are obligated to that may not be written down in the law. We're a commission of ethics, not just of law. So I wonder if -- because any consent in writing that isn't extraordinarily specified is going to open itself up to the misuse of data as potentially incriminating as, you know, your whole genome sequence might be.

MS. SUTER: Right and I think this is why I really emphasize the notion of privacy as trust-based, that there are these obligations to think about what the data is being collected for, what the uses are and what protections. So the inappropriate re-identification or public disclosures would be something that as you say we can't come up with every possible condition in advance, but that we have to as a community think about those kinds of moral obligations that the people who are entrusted with the information and data have to care about and protect.

DR. GUTMANN: Thanks. I think it is important to see the connection

between privacy and control. That is privacy protects -- our sense of privacy is also a sense of control of not only who uses but how information about us is used, even when it won't harm us. There is -- we have an article in our readings by James Rachel on the value of privacy which points out something that is so obvious but so easily overlooked, that a married couple may have nothing embarrassing about their sex life but they still don't want people observing it. At least most of us don't.

(Laughter)

DR. GUTMANN: And if they do they can allow it, right? And the same thing is true for dinner conversations at restaurants where you're doing it, not shouting at the top of your lungs. So, there is -- you don't have to believe in whole genome sequencing being exceptional, you just have to believe what's true which is there's a lot of information that can be widely shared. And it may not even harm you in any strong way, it just may be absolutely ethically inappropriate for people to use it in ways for which it wasn't intended, that you didn't intend it to be used.

MS. SUTER: Right.

DR. GUTMANN: I just wanted to make -- I think your -- both of you, what you spoke is totally consistent with that, but it's just, it's a very simple point that's very easily overlooked that underlies a lot of people's unease with the idea that it would just be out there in the world used by anybody in any way they wanted to just because you gave consent for some organization to use it in ways that you assumed would be ethically sound.

Okay, that's my, just my takeaway from some of what you've said. Jim and John I have on the list and I'm looking. Okay.

DR. WAGNER: Looks like everybody's on the list.

Well first of all, Professor Suter and Dr. Sweeney, thank you so much for the presentation. This is billed as a technology section and Dr. Sweeney, you mentioned something about technology about which I'm not familiar or aware, and it's this concept of limited data-sharing. Is this perhaps a possible solution for us? If limited data-sharing as you described it briefly, the ability to ensure that the analysis of data can be processed externally but the data itself cannot be leaked out and moved onto other folks surreptitiously, is that technology really available and does it do what I'm saying? And presumably it means that down the road I could say okay, that's enough direct sharing of my data and I could turn it off?

DR. SWEENEY: Yes. So this is a really interesting observation that you're making because -- so I'm going to just back up a few levels before I answer that question. When we look at Facebook and Google who basically monetized the giving away of personal information and therefore really pushed forward in the public view the concepts of open data and open consent all get rooted in this idea that so much data is being given away freely and look, after all I get these services.

But then when you go as a researcher and you go to Google and you go to Facebook and you say you've got great data, can we do something to -- we could learn something for society from this, the answer is oh, we couldn't possibly give you access to this data. So the data gets blocked down into these silos and all of a sudden it becomes the asset of this company and as an asset of the company they have entrenched interests in keeping the data very much where it is and very secluded.

And we're starting to see that pattern also happening in health data. As more and more health data outside the use of the patient is being monetized people build up these large databases that are unaware by the public but for which no one -- researchers can't get access and so forth.

And so these kinds of silos basically are bringing forward these ideas that they don't even -- even if I want to produce a new revenue stream and I've got a data set and you've got a data set. I don't want to share my whole data set with you but together we want -- this is one of the best places where we're finding technological desire to keep the data in place and actually do computations across the data.

And being able to do that in a cloud computing environment makes it very, very fast as well while each of the data sets can be encrypted and walled and things like that. So this is one of the places where it's exactly these -- it's interesting because it's exactly the silos who push us towards open data and public are pushing us towards these other technologies.

DR. WAGNER: If we were going to make reference to this kind of technology and mechanisms for permissive use of data without release of the whole data sets what's the current technology? Is it this direct -- terminology. What's the correct use of terminology rather?

DR. SWEENEY: So there are three ways depending on what the level is. One is -- goes under a general umbrella called multi-party computation is a technical umbrella which has sped up recently a lot by -- so the example I gave in my talk was PrivaMix is a good example of that, a real world use of that.

The second level at which this is happening which I also talked about in he talk was -- and also by the way NIH I think is also trying to bring forward an environment like that too. I can look up its name.

But the other one that came up in the talk is again this MyDataCan environment which says suppose we keep you in the silo and we're just going to control in very refined increments whatever could ever leak out. So a lot of times what happens in these privacy risk issues is that I have -- the only reason I have to have sensitive data is because I want to link it or something like that, but the result turns out to be fine. Well, if you have MyDataCan or some large data set like that you can do computations on the data and the computations that you take out of the data, the correlations as it would be in genomic data turn out not to be very sensitive, but the data on which the correlations are based could be sensitive. And so this is the other way we see this.

DR. WAGNER: And you would call that?

DR. SWEENEY: Well, we might call it data-mining, we might call it primarily data-mining. And we see that a lot happening as well.

DR. GUTMANN: John?

DR. ARRAS: Thank you very much for those two presentations. So, Dr. Sweeney, with regard to Data to Science, okay, so you're arguing that at the person's demise all of this information gets accumulated, okay?

Two issues. One is this assumes that we've already solved the problem of who owns the data, right? The individual or the universities or the corporations that, say, sequence the genome, the genomic information. So, do these living labs have a kind of

policy component that grapples with those kinds of questions? Or are you arguing that the individual should properly be viewed as the owner of, say, you know, that sequencing data?

The other question that I have for you is I guess a kind of naive question. What's so special about death? You know, I mean if we have adequate privacy protections, right, why should we build a living lab around, you know, a cutoff point of death?

DR. SWEENEY: So, to answer the first question about data ownership. So, in the models of the living labs that I gave the idea is that there exists a rich copy of data that belongs to the subject of the data to the extent the subject can actually get the data. And we increasingly see like in the high tech act we see these provisions to allow individuals to get access to data that's often caught in these silos. And there seems to be movement more and more in policy to help individuals get access to their copy of the data. And even in cases where my phone company may not make my text messages and so forth, there are apps that will actually capture them for you and store them in a repository like MyDataCan.

And so -- and many people would argue, some of my colleagues at MIT have shown what happens when you combine data across silos, that you can predict disease, you can improve the efficacy of drugs and things like that. And so we can show that people can get a better idea -- a better life. So the idea is that the copy that the individual assembles is their copy and they have ownership of that copy. It doesn't place any issue on the copies that are elsewhere out there. So it's not ownership of all this data about me is mine, it's just only that this copy that I put some energy in and assembled is my copy. And we believe that it's a richer copy so therefore we can do other things with it and that's sort of its strategic advantage.

The Blue Button Campaign is another example where we see individuals being able to get, for example in the case of the Blue Button their medical records. So increasingly more and more facilities, health insurance facilities and hospitals and pharmacies and so forth allow you to go to a portal and download your data. So this would be an example of being able to achieve that data.

You asked what's important about death. So one of the things -- we have a different project that deals with privacy assessment and tagging. And so normally when you think of the harms, and we tend to focus on economic harms from privacy. These economic harms, sort of the most egregious tend to be criminal. So that I could end up finding myself in a criminal situation. And that in the case of genomic data is if you have a large repository of genomic data, some incident happens in the population for which there's a genetic sample left at the scene and now law enforcement says aha, you hospital, you have this big one. They get a search warrant and so forth. So there are these criminal kinds of ideas.

There are civil and economic problems, you know. There was a study many years ago where Fortune 500 companies admitted to making hiring, firing and promotion decisions based on medical data. And there have been other reports of problems happening, but notice you don't see a lot of them. And the reason you don't see a lot of them isn't because they're not necessarily happening, it's because the data-sharing is hidden. And the more the money gets into the data, the further -- the more hidden it becomes. And so these harms are out there.

But one thing -- so MyDataCan allows you to collect the data while you're

alive and you can participate in research. There are mechanisms for doing that. But the particular focus on death is usually that in general we say that a person's privacy interest dies with them. And when we look at those kinds of economic harms that we just talked about they tend to not be possible once the person dies.

DR. GUTMANN: The reputational harm.

DR. SWEENEY: I know. I know. I'm among ethicists. I'm a computer scientist and I have no ethical position.

(Laughter)

DR. SWEENEY: So let me just explain.

DR. GUTMANN: That's not a pure -- I just want to defend or not defense of ethicists. It's not -- it's an empirical statement that people's reputational interests continue. People care about their reputations after death.

DR. SWEENEY: I agree.

DR. GUTMANN: What value you put on it may be arguable, but that is --

DR. SWEENEY: I agree. And I'm just saying that the reason -- the focus on the reason for death is that the economic harms tend to go away and so as a result of that, and also for other reasons it can be easier to acquire the data. And that was why the focus on death. And I just want to repeat I'm not an ethicist.

(Laughter)

DR. GUTMANN: Thank you for being very forthcoming. We're just trying to refine our own thinking on this. Nelson.

DR. MICHAEL: Thanks to both of you for those presentations. So my

question primarily is for you, Dr. Sweeney. So you described what sounds like to me an open laboratory where members can join and essentially do experiments within the context of this laboratory. Is the ultimate intent to learn to massage this new world so that members of the public or interested parties can become part of this open laboratory and eventually influence decisions about how this technology should be used, how privacy can be preserved.

And if that's the case, if the ultimate goal of this is to influence practice and to determine what best practice is and to explore ethical dimensions, aren't you by the very nature of what you're doing, doing research? And therefore do you believe that by doing research you need to be under the kinds of usual constraints that universities would regulate research for?

DR. SWEENEY: So first of all, by "living lab" what we mean by that is that we want to create an exciting environment that people want to come and visit because it just makes their life -- it's fun and they can do things and they learn things about themselves that they wouldn't know before. And that's possible because we've assembled all of this different kind of data about them and we have these apps that are helping them learn more and live better.

And the reason we call that a living lab as opposed to going out and doing it as a company is because we purposely are doing this also in a way that has certain privacy guarantees and certain privacy assertions and certain privacy questions.

So the first questions is, you know, will a million people come forward in 5 years. Will people engage in this? Will we be able to deliver on that excitement about what

your data can do for you?

And then if they do it's a living lab because they will begin to shape what the experience is there. It's not that we're doing experiments on them as much as they become participants that actually will guide, well, we want this in our platform, we want this kind of thing and so forth.

In 5 years we might be able to do an interesting experiment, an interesting study that says what is the maximal social utility with respect to privacy in this environment. You know, she mentioned selection bias and things like that. We believe that you won't see the selection bias that we've historically met because it's so multifaceted. The people choosing to participate or not participate in one experiment won't be an issue because there are just so many others and other things that are drawn to it.

So we believe those kinds of things we'll be able to show and prove and that the maximal social utility will be better under this kind of regime because individuals can make fine-tuned decisions that you can't make in a draconian policy than say a regime like HIPAA, for example. That would be one of the questions.

And then there's the point that you made, ultimately this is a research environment. Everything that we do for intervention requires us to go back to the IRB if we were actually intervening. But the structure itself doesn't require intervention in general because we're just trying to build an exciting platform.

DR. GUTMANN: Thank you. We have a lot of questions so let's be as brief as you can be. I will just cut us off at the time. Dan.

DR. SULMASY: Thanks. Another question for Dr. Sweeney. I'm trying

to understand a bit more about the sort of protections that come from something like MyDataCan.org. Because it seems to me if I'm understanding you correctly that owning a copy doesn't mean that we're taking the information away from the control of Facebook or whoever, we just have a copy of it. And we're creating a copy then in which somebody -- where you could put it all together in one place, you know, my DNA sequence, my medical information, my bank accounts, my magazine subscriptions, my grocery bills, where I've lived, GPS information, Facebook account. And so we're not getting any more protection by taking it away from the commercial interest that might sell it, and we're putting all of this together in one place which for me seems like the biggest sort of threat that all of that could be cross-linked together.

And I don't know how many, you know, levels of encryption you need to keep somebody from hacking into that, but that seems to me to be the sort of -- a very deep threat to people's privacy when all of that can be put together.

DR. SWEENEY: Well, I mean other than to say that our inability -- that every person has a private key and no two people have the same key so that even if the data were broken into and even if you had Harvard's copy of the key it would be impractical for you to figure out what these various pieces of data might be. So we feel very confident on that end of it, on the data storage against the adversary who breaks in, that kind of thing.

The issue in all of these things continues to be the data that's given away. And being our ability to control that through the apps. The fact that the data is assembled by the individual and under the individual's control and our use of technology to enforce that we feel pretty good about. Making sure the data-sharing is transparent and that no app

is leaking data, grabbing data it's not supposed to, things like that, is also extremely important.

DR. GUTMANN: I have two questions from members of our audience that are similar and directed to Dr. Suter. One comes from Bartha Knoppers. Correct me, Bartha, if I'm wrong here in reading your handwriting but it says, "Are opt-out clauses in research, are they unworkable unless" then I can't read. Well, the basic question is are opt-out clauses unworkable. And Bartha, you can expand on this for a minute but I want to add Steven Sherry.

Steven who is a senior scientist at NIH has a similar, an intersecting question. So again, Dr. Suter, in the future having studies with tens of thousands of participants are there a finite and manageable number of categories of opt-out uses? If not, the long list of opt-out choices can quickly multiply and reduce the number of effectively consented individuals to a level where the study is underpowered. So they're both questions of whether opt-outs are manageable. Bartha, have I captured yours? Do you want to add anything to that?

DR. KNOPPERS: That's fine.

DR. GUTMANN: Okay, great. Thanks. Dr. Suter.

MS. SUTER: So I think that's a fair point, that if you start to make it very, very specific about all the various things one opts out it starts to look a little bit like the reverse of not true informed consent but specific consent for different kinds of studies. So I think the idea is to focus on categories to try to limit it.

It does lead to some impracticalities and I think that's sort of the

compromise of research. It's going to be a little bit more difficult. Although I'm not an expert on how data tracks the way Dr. Sweeney is, but I'm guessing that there could easily be things that tie to the samples so that there could be a fairly easy weeding out of things. This is my hope, that technology would offer that sort of possibility. But that it has to be sort of a middle of the road, not a complete blanket consent where you consent to everything, some opt-out, but limiting categories so it's not completely limiting in the abilities to do research or time-consuming. So a compromise.

DR. GUTMANN: So I have four Commission members on my list. I'm going to ask them to ask their questions and then ask you to answer them in a group. Christine, Nita, Anita, Dan -- oh no, Dan did. Raju. Raju who said by sitting next to me he would be overlooked.

(Laughter)

DR. GUTMANN: But Raju you get the cleanup question. Okay, Christine, Nita, Anita, Raju.

DR. GRADY: Thank you and I echo everybody else's thanks for your presentations.

I want to follow up a little bit on this notion of trust because I think that I agree that this is -- a lot of it depends on the notion of trust. And I worry a little bit that consent certainly in the way we usually think about it but even in the ways we're talking about it is not -- can't bear the weight of all of what we need to communicate with the public in terms of expectations upon which they base their trust.

So I wondered if you thought about first of all whether or not consent

really is where we should be focusing and if not what are the other things we should be focusing on to help the public have different expectations upon which they will then base their trust.

And I think it's really interesting to me and some of the control issues happen this way too that we somehow trust or don't think about it, you know, Google or Facebook, but we don't trust the researcher or the doctor who's trying to help us. And so how do we --

MS. SUTER: I don't trust Facebook.

(Laughter)

DR. GRADY: Well, some people do. So, I mean it's just sort of interesting. How do we build those expectations of trust? That's the question. And I'd love to hear from both of you if you have time.

DR. GUTMANN: You have to wait for the other questions first.

DR. FARAHANY: Again, thank you. This has been an incredibly enlightening conversation. There's two different issues that I've heard echoed throughout that I wanted to focus on a bit. The first is control over data. The second is the reasons why somebody might actually have some control over genetic information in particular. And Professor Suter, you mentioned that you're not a genetic exceptionalist, you recognize that there may nevertheless be some sensitive uses or sensitive implications of genetic information.

But in thinking about both control and particular reasons why you might have control over genetic information I wonder if you could be more specific as to whether

or not you think you should have control over any use of genetic information or only particular uses.

And when Professor Suter mentioned Fourth Amendment jurisprudence which is more limited for a particular purpose in those cases they've all dealt with just identifying uses of genetic information. Most of the courts have gone on to say if it were used for other purposes then there may be privacy implications but that individuals have no reasonable expectation of privacy in identifying information. And that genetic information is no different than any other type of identifying information from being able to see your face, being able to see health issues that you have. Even though I can't read your health records if it was visible to me and I could tell you wouldn't have a privacy interest in that.

So in thinking about control I'm wondering if it really is unfettered control or much more specific about control over particular uses or applications of genetic information. Thanks.

DR. GUTMANN: Anita.

DR. ALLEN: Thank you, Dr. Sweeney. You're sort of famous for doubting how successfully we can keep data de-identified. And I think many people associate you with the view that it's way too easy to re-identify previously de-identified information. And in the genomics context, in the medical context in general we talk a lot about oh, the data is de-identified, therefore it doesn't matter if it gets shared.

Could you just comment from your expertise on how much we should worry about the potential for re-identifying de-identified information? In the sort of INS context but in a more broad context.

And then I also was very interested in this whole question about controlling data and its association with privacy because I've always had two problems with the idea that controlling data is what privacy is all about. And one concern was raised today already. If the data is controlled but is simply a copy of the data and lots of other people have their own copies where's the privacy in that?

The second point is that if the data is controlled by you but you choose to continually share the data where's the privacy in that? Unless we think that privacy only means having the ability to control the data but always publicizing it. So those are just my two comments -- questions.

DR. GUTMANN: Raju.

DR. KUCHERLAPATI: Yes, thank you very much for your comments. My question is somewhat similar to what Anita was asking. And I think about these privacy issues not just in terms of research but also sequencing that is done for clinical purposes. Because I believe that as we move into the future there are going to be a greater number of individuals who want to have their DNA sequenced for clinical purposes.

So the question is that the concern about this obviously is that some information that you'd want to keep private if it becomes public that it will cause harm. For example, like Amy talked about the fact that one could go to a physician and the physician would diagnose you to have a psychiatric illness, for example. And that is done today and that is part of the medical records. Or you could have genomic information, whole genome information and somebody would be able to glean from that information that a person has a susceptibility to get a psychiatric illness. I was wondering as to whether you consider

whether those two things are different.

And I also want to ask the question obviously currently we have privacy restrictions about how to deal with clinical data such as those that are, you know, managed by HIPAA and whether we need to have new regulations to think about genomic information, and whether that is intrinsically different than having clinical information like the diagnosis by the physician.

DR. GUTMANN: You're on. You can divide up your answers, just -- I know you can't be comprehensive but if you would between the two of you take a swipe at answers to the four questions.

DR. SWEENEY: Okay. On the question of trust I think it's a fair point that consent can't do it all which is why I started with the notion that there are certain things we have to do to build the trust. Try to ensure that the research is done for the public good.

I think educating the public is a useful approach although I'm skeptical about how much you can do in that. I mean, as a genetic counselor I found one-on-one sessions only educated people so much. But still we can educate the people far better than we have done and I think that would be a place to start.

And I think the other point of it is to really emphasize that the people to whom the information is disclosed have these obligations as trustees. And I think that has to be built into the standards of practice and very clear about sort of how far can we really go with the research.

The question about unfettered control, or whether it was unfettered control or control over specific uses. I think what I'm trying to deal with is hoping that the trust

within the doctor-patient relationship and the research setting will offer many of the protections that Dr. Sweeney is working on to try to keep the data secure. But that we try to have control over all sorts of uses outside of that that would be deeply problematic.

The fact that people could get all sorts of information right now without prohibiting -- violating any laws in many states. It's true that the Fourth Amendment cases focus on identifying information, but crime labs are able to retain the samples for a long time. So although theoretically it's only about identification they still have access to the information. And so clear prohibitions in most states there are that they don't go beyond the identifying information.

And then the last question was about control about privacy. I think that you asked a question, I'm not quite sure how you worded it. You said if some people want to share the information is that privacy. So sort of letting people waive their right to privacy seems to be your question. And I guess I would be okay with that.

I think that this is part of the autonomy interest that underlie privacy, for people to have the ability to decide how private they want to be. The couple that talks about their sex life, right? That's a choice people make.

In the genetics context though it does potentially have implications for other people who may want the information private like family members. And so that makes it a little bit different. Or the wife who talks about it and the husband who doesn't want her to. So, I don't think I addressed it all, but.

DR. GUTMANN: Not a very happy marriage.

(Laughter)

DR. SWEENEY: One big chunk of putting them all together. There were a lot of questions there around the control issues. And that also linked back to the fine-grained consent.

So the motives behind things like the kinds of systems that I talked about today is trying to say that technology offers the opportunity for individuals to have input into data-sharing arrangements unlike they've ever had before. So the idea of informed consent in these normal consent models, this is the place actually where technology can come into vogue. Because it can keep provenance of what was agreed to and under what conditions, and it can offer a fine-grained notion of permissions. And so that's the high level of that.

The control over their own use adds to transparency which improves satisfaction and trust. The idea of asking, there have been numerous surveys that people just want to be asked and if they are asked they will normally often say yes.

And I do think that that's -- the Personal Genome Project should always be in the back of all of our minds as an example. They have over 1,000 people who say the most intimate details of depression, abortions, medical conditions and so forth. It's all right there on the internet. Just anybody can look at it. It's pretty amazing.

The idea of why would you then want to assemble all the data under an individual's control but you don't have anything to say about the others is because we actually are doing exactly what Facebook and Google did to us to lead us to open data. If we were successful and many Americans march to the tune of something like MyDataCan then it will actually push back on those other data sets, that they don't have the same

permissions, they don't have the same level of trust, they don't have the same level of satisfaction.

And then not only does it produce a different value proposition that the quality of the data that's under the individual's control is better, but that also individuals trust the use of the data under that kind of system better, and therefore will have policy implications downstream. Research questions, part of our research questions. We don't know that that's true.

HIPAA, the last question there was HIPAA. I thought the observation of a clinical data is absolutely right. Whatever we think about genomic data today, in 5 years it would probably be most of what we have to think about genomic data today will be a part of the clinical record. Between various tests and so forth that are done it will be there. It's under HIPAA but HIPAA is not complete coverage. Multiple organizations get exactly the same copy of the data that is covered under HIPAA so this group has to play under one set of rules.

This group has the same copy of data and they don't have to play under those rules which then leads to the re-identification question that Professor Allen brought up and that is that that's actually why the re-identifications are so much easier in health data, because you've got regimes who have the same data but are playing by releasing the data in different rules. And so even -- if you do various redactions here you can't say what hospital discharge data or all claims databases or HIEs or whatever they're going to release the data under will therefore make it able to be linked to. And since we can't control all of them we constantly leak the risk of re-identifications.

DR. GUTMANN: I want to thank you very much since you couldn't control our questions but you did a great job in handling them. So on all our behalfs let's thank Dr. Suter and Dr. Sweeney.

(Applause)